主观社会阶层正向预测利他性惩罚*

陈思静[1] 杨莎莎[2] 汪 昊[1] 万丰华[1]

([1]浙江科技学院经济与管理学院, 杭州 310023) ([2]上海大学经济学院, 上海 200444)

摘要 利他性惩罚是指个体自行承担成本来惩罚违规者的行为，它受到社会阶层的影响。研究 1 利用 2013 年中国综合社会调查数据发现阶层显著正向预测利他性惩罚。研究 2（N = 450）基于真实生活事件的调查表明，惩罚成本调节了阶层对惩罚的影响：在成本较高的直接惩罚中，阶层正向预测惩罚；而在成本较低的间接惩罚中，这种作用不再显著。研究 3（N = 232）通过操纵阶层与成本进一步证实惩罚成本的调节作用：高阶层者比低阶层者更有可能做出利他性惩罚，但两者的差距在高成本条件下更突出。研究 4（N = 125）综合考察了阶层影响惩罚的心理机制，多层线性分析显示：惩罚成本较低时，阶层通过公正世界信念来间接影响惩罚，而成本较高时，阶层直接正向影响了惩罚。上述结果意味着利他性惩罚受到个体社会阶层的影响，同时也在一定程度上说明在利他性惩罚中基于成本-收益的策略性考虑并非完全缺席。

关键词 利他性惩罚，社会阶层，公正世界信念，惩罚成本

分类号 B849: C91

1 引言

利他性惩罚（altruistic punishment）是 Fehr 和 Gächter（2002）用来解释人类非亲缘个体间大规模合作的一个重要概念。在这类惩罚中，惩罚者承担成本来惩罚违反社会规范的个体，但这并不会给惩罚者带来直接利益。利他性惩罚具有三个特征：1）惩罚者自行承担惩罚成本；2）惩罚给违规者造成了损失；3）惩罚维系了社会规范（李佳等, 2012）。利他性惩罚对合作的促进作用得到了大量研究的支持（Balliet et al., 2011）。利他性惩罚的研究多在实验室环境下展开，而为了得出清晰可靠的因果结论，参与实验的被试往往是同质的、原子化的个体（Gowdy, 2004; Manner & Gowdy, 2010）。但个体并不是生活在真空中的原子，恰恰相反，个体的先天遗传和后天生活经历导致个体之间存在显著差异，其中一个重要差异是源于阶层的差异（杨沈龙等, 印刷中）。陈云松和范晓光（2016）指出，大部分人都具有阶层意识，并

收稿日期：2022-4-5

* 国家自然科学基金项目(71701185), 浙江省哲学社会科学规划课题（22NDJC114YB）资助。

通信作者：陈思静, E-mail: chensijing@zust.edu.cn; 杨莎莎, E-mail: sanqijiuba@163.com

且倾向于将自身定位于某个阶层中来认知和行动，而来自不同国家的研究者指出当代社会的结构越来越表现为阶层化（李路路, 2012; Law & Law, 2006），即不同的个体分属不同社会阶层，且阶层边界越来越清晰。近年来全球范围内贫富差距的急速扩大更是强化了这种趋势，贫富差距导致人们更加倾向于使用阶层这个维度进行社会归类和认同，并表现出基于特定阶层的认知与行为模式（Tanjitpiyanond et al, 2022）。因此，一个富有意义但在惩罚领域尚未受到足够关注的问题是：社会阶层如何影响利他性惩罚？本文通过四个研究较为完整地考察了这一问题，并给出了初步的答案。将社会阶层引入利他性惩罚一方面可以拓展社会阶层的研究领域，从而丰富社会阶层这一概念的理论内涵；另一方面则可以更好地理解源于阶层的个体差异如何影响了利他性惩罚，从而提升惩罚研究的生态效度。

1.1 社会阶层的社会认知理论

基于社会认知的视角，社会阶层通常被定义为在社会层次结构中处于不同地位的群体，这些群体占有不同的社会资源并表现出不同的自我概念与社会认知（Kraus et al., 2012）。社会阶层可以通过社会资源的客观占有来评定，如收入、教育和职业（Stephens et al., 2014）；也可以通过自我估计来评定（Christie & Barling, 2009），前者被称为客观社会阶层（objective social class）或社会经济地位（socioeconomic status），而后者被称为主观社会阶层（subjective social class）（胡小勇等, 2014）。社会阶层对个体的方方面面产生了重要影响（胡小勇等, 2014; Kraus et al., 2010; Stephens et al., 2014），如自我概念：低/高阶层者分别发展出互依（communal self-concept）或独立的自我概念（agentic self-concept）（Stephens et al., 2011）；社会知觉：知觉他人时，高低阶层者均觉得温暖（warmth）比能力（competence）更重要，但这一效应在低阶层者中更明显；而知觉自身时，低阶层者赋予了温暖更高的重要性，而高阶层者刚好相反（韦庆旺等, 2018）；人际关系：高阶层者偏好交换的人际关系（exchange relationship），而互依的人际关系（communal relationship）则受到低阶层者的青睐（Kraus & Keltner, 2009）；身心健康：相比高阶层者，低阶层者更易受到身心疾病的困扰，并且死亡率也更高（Elo, 2009）；利他行为：相较于高阶层者，低阶层者更愿意实施利他行为（Piff et al., 2010; Piff et al., 2012; Piff & Robinson, 2017; Stellar et al., 2012），尽管有研究者指出这一效应受到诸多因素的调节，包括被试年龄（Benenson et al., 2007）、行为对象（Kuang et al., 2021）、行为是否公开（Kraus & Callaghan, 2016）、群体内的不平等是否突出（Côté et al., 2015）以及利他行为的类型（Penner, 2005）等。上述文献回顾表明社会阶层会影响人们的自我概念、人际关系以及包括利他行为在内的种种现象，而鉴于利他性惩罚属于利他行为的一种，我们推测利他性惩罚同样受到社

会阶层的影响，但遗憾的是，目前尚无研究系统地探索两者间的关系，为弥补这一空白，本文拟从社会阶层这一新视角来检验利他性惩罚在人群中的分布、心理机制与影响因素。

1.2 社会阶层对利他性惩罚的影响：基于资源与动机因素

尽管尚无研究系统地考察社会阶层对利他性惩罚的影响，但有若干间接证据可以帮助我们推测两者间的关系。从客观条件来讲，利他性惩罚作为一种高成本信号（costly signal）本身需要惩罚者承担相当的代价（Jordan et al., 2016a; Nelissen, 2008），而高低阶层之间的重要差别即是对社会资源占有的不同（Stephens et al., 2014），高阶层拥有更多的资源，因此在其他因素相同的情况下，我们可以合理地推测高阶层者更有条件做出利他性惩罚。其次，来自公正动机（justice motive）的研究也显示，高阶层者拥有更高的公正世界信念（belief in a just world, BJW）（Furnham & Procter, 1989），而公正世界信念可以正向预测个体的惩罚态度（Bègue & Bastounis, 2003）或惩罚行为（杨莎莎, 陈思静, 2022），因此从动机上来讲，高阶层者似乎也更有可能来惩罚违规者。此外，也存在一些实验证据来支持上述推测，如 Ding 等（2017）的最后通牒博弈实验显示，富裕的被试更容易拒绝不公平的分配方案，而最后通牒博弈中的拒绝行为通常被认为是一种利他性惩罚，因为个体通过付出成本减少了违规者的收益，并在一定程度上维护了公平规范（李佳等, 2012）。基于上述证据，我们提出第一个研究假设：

假设 1：社会阶层正向预测利他性惩罚：个体的阶层越高，越有可能做出利他性惩罚。

1.3 惩罚成本与公正世界信念的视角：一个条件过程模型

从经济学视角出发的惩罚研究往往关注基于成本-收益的策略性考虑在利他性惩罚中的作用，如范良聪等（2013）以及陈世平和薄欣（2016）发现，利他性惩罚同样遵循经济学中的需求曲线，即商品价格（惩罚成本）越高，需求量（惩罚行为）越低，从这个意义上来说，利他性惩罚似乎和普通商品的购买行为并无二致。陈思静等（2020）的研究也显示，变更惩罚成本的形式并不会影响上述结论，只会导致惩罚与成本之间的关系从一条需求曲线转向另一条需求曲线。总的来说，随着惩罚成本的上升，惩罚频率或强度总体趋于下降（Aharoni et al., 2019; Anderson & Putterman, 2006）。上述发现说明基于成本-收益的策略性动机在利他性惩罚中确实存在。更为重要的是，不同社会阶层的个体拥有不同的资源（Drentea, 2000; Oakes & Rossi, 2003），这导致他们对成本的敏感性也有所不同（Paulsen & John, 2002），高阶层对成本的敏感性更低，因此，尽管惩罚成本的升高对高低阶层的惩罚行为均有抑制作用，但这种抑制作用对低阶层应该更为明显。由此，我们提出如下假设：

假设 2：惩罚成本调节了社会阶层对利他性惩罚的影响：总体上高阶层者比低阶层者更有可能做出利他性惩罚，但两者的差距在高成本条件下更突出。

不同于基于成本-收益的研究，基于惩罚动机的研究证实了利他性惩罚中存在非策略性动机，因为利他性惩罚在一定程度上由公平原则驱动，这意味着被试在实施利他性惩罚时很可能是为了维护某种内心的信念（Falk et al., 2015）。有研究者运用 Lerner（1965）的公正世界信念来解释上述发现：个体需要相信自己所生活的世界是公平的，因此愿意付出代价来惩罚破坏公正世界的违规者（Strelan et al., 2017）。此外，Furnham 和 Procter（1989）发现，社会阶层与公正世界信念之间存在正相关，而公正世界信念又可以正向预测个体的惩罚态度（Bègue & Bastounis, 2003）或惩罚行为（杨莎莎, 陈思静, 2022），且有证据表明公正世界信念中介了社会阶层与网络利他行为之间的关系（郑显亮等, 2021），我们推测在社会阶层与利他性惩罚之间公正世界信念也起到类似的作用。需要说明的是，公正世界信念包括个人公正世界信念（personal belief in a just world）和一般公正世界信念（general belief in a just world）（Wu et al., 2011），并且后者通常和个人的惩罚态度（Bègue & Bastounis, 2003）或惩罚行为有密切关联（杨莎莎, 陈思静, 2022），因此本文主要考察一般公正世界信念的中介作用（为方便行文，如无特别说明，后文中公正世界信念均指一般公正世界信念）。由此提出最后一个假设：

假设 3：一般公正世界信念中介了社会阶层与利他性惩罚之间的关系，具体而言，社会阶层正向预测一般公正世界信念水平，而后者又进一步正向影响了利他性惩罚。

最后，我们结合惩罚成本和惩罚动机内外两种视角，在检验假设 3 的基础上通过引入惩罚成本这一调节变量，提出了一个社会阶层影响利他性惩罚的条件过程模型（图 1），这样做有两个目的：1）综合考察社会阶层影响利他性惩罚的心理机制与边界条件，可以为本文的研究主题提出一个初步的解释框架；2）通过检验该模型，我们也可为理解利他性惩罚中策略性动机和非策略性动机的关系提供一种新的思路。需要说明的是，社会阶层包括客观社会阶层和主观社会阶层，且有研究者发现两者之间存在中等程度的相关（Adler et al., 2000），对个体具有独立的影响（Manstead, 2018），同时有研究表明后者比前者能更好地预测个体的行为（Cohen et al., 2008; Wolff et al., 2010），因此，本文一方面同时考察客观阶层与主观阶层对利他性惩罚的影响：参考以往研究（e.g., Kraus & Tan, 2015），我们选用收入和教育作为客观阶层的两个指标；另一方面，本文在讨论结果时，主要关注主观阶层对利他性惩罚的影响，如无特殊说明，本文中社会阶层均指主观社会阶层。除有特别说明，本文数据分析均通过 SPSS 25.0 实现。

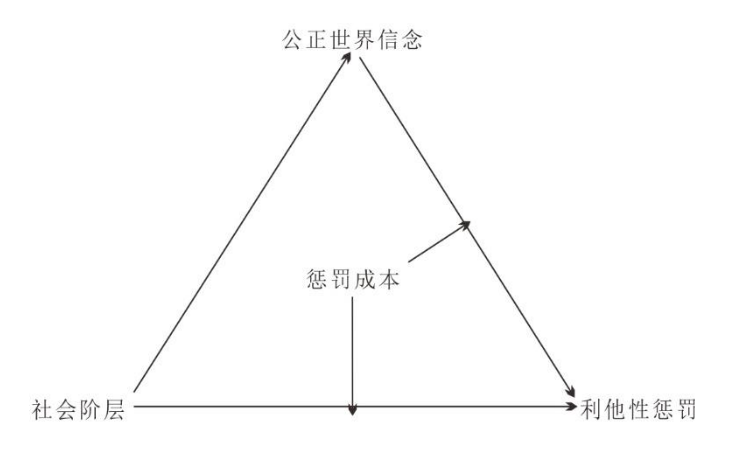


图 1 社会阶层影响利他性惩罚的条件过程模型

2 研究 1

2.1 数据来源与变量选取

研究 1 利用中国综合社会调查数据（CGSS2013）（中国人民大学中国调查与数据中心, 2015）检验了社会阶层与利他性惩罚之间的关系，我们选取了以下变量来验证假设 1。

主观社会阶层 在研究 1 中，预测变量为主观社会阶层。我们选择了被试对题项 A43a 的回答作为主观社会阶层的指标。在该题项中，作者采用了 Adler 等（2000）所开发的方法，向被试展示了共含有 10 个层次的阶梯，10 代表社会最高层，而 1 代表社会最底层，被试从 1~10 中选择一个整数来代表自己所处的阶层，而回答为“−3=拒绝回答；−2=不知道；−1=不适用”的被试则被排除。

利他性惩罚 在研究 1 中结果变量为被试的利他性惩罚。CSSS2013 中有两个题项和利他性惩罚有关，分别为举报社会问题（D13：如果您所在的单位有一项举措可以提高集体福利并使您个人得到利益，但会造成环境污染或社会公害，您会举报吗？−3=拒绝回答；−2=不知道；−1=不适用；1=会；2=不会）和反抗上司不端行为（D23：假设您的上司或老板是外国人，如果他侮辱了中国，但抗争会产生不利于自己的后果，您会选择：−3=拒绝回答；−2=不知道；−1=不适用；1=当面抗议；2=保持沉默；3=暗地里报复；4=以屈求伸，背后骂几句就行了；5=无所谓）。基于 Molho 等（2020）方法，将上述两题的选项转换为二分变量：1）对于举报社会问题，我们排除了回答为−3、−2 或−1 的被试，并给阴性事件赋予了较低的值，因此转换后结果为：0=不会，1=会。2）对于反抗上司不端行为，同样排除了回答为−3、−2 或−1 的被试，并将选项 1 和 3 合并为惩罚，赋值 1，选项 2、4 和 5 合并为不惩罚，赋值 0。

控制变量 被试的性别（A2：1=男；2=女）、年龄（A3a）、最高教育程度（A7a：1=没有受过任何教育；13=研究生及以上）和年收入（A8a：您个人去年[2012]全年的总收入是多少？）为控制变量。在性别题项中回答为“−3=拒绝回答；−2=不知道；−1=不适用”的被试、在教育题项中回答为“14=其他”以及在收入题项中回答为“9999997=不适用；9999998=不知道；9999999=拒绝回答”的被试均被排除。由于和其他变量相比收入的绝对数字较大，对收入进行了对数化处理，函数选择了 ln（1+年收入），因为部分被试报告其年收入为 0。

对样本进行上述筛选后，我们最终得到了 4978 份有效数据。被试平均年龄为 48.57 ± 16.04 岁，女性占比 48.83%。

2.2 结果与讨论

各变量的描述统计和相关系数如表 1 所示。

表 1 研究 1 变量描述性统计与相关系数

变量	*M*	*SD*	1	2	3	4	5	6
1 主观社会阶层	4.33	1.67						
2 惩罚 D13	$P_{惩罚}=0.57$		0.06***					
3 惩罚 D23	$P_{惩罚}=0.61$		0.05***	0.29***				
4 性别	$P_{女性}=0.49$		0.02	−0.06***	−0.09***			
5 教育程度	4.91	3.02	0.21***	0.02*	0.06***	−0.13***		
6 年龄	48.57	16.04	−0.07***	−0.02	−0.01	−0.04**	−0.46***	
7 年收入	8.49	3.30	0.12***	0.04**	0.06***	−0.26***	0.47***	−0.03*

注：惩罚 D13、惩罚 D23 和性别均为二分类变量，故此处呈现占比情况；性别处呈现的是女性在总体样本中的占比；教育程度为 1 表示没有受过任何教育，13 表示研究生及以上；年收入已经经过对数化处理；$^{*}p < 0.05$，$^{**}p < 0.01$，$^{***}p < 0.001$。

以阶层为预测变量、利他性惩罚（举报社会问题）为结果变量做二元 Logistic 回归（$N = 4921$），模型拟合度良好（$\chi^2(5, 4921) = 38.93, p < 0.001$; Hosmer-Lemeshow 检验不显著: $p = 0.597$）。回归分析结果显示：年龄（$B = −0.003$, $Wald = 2.80$, $OR = 1.00$, 95% CI [0.99, 1.001], $p = 0.094$）、收入（$B = 0.01$, $Wald = 1.45$, $OR = 1.01$, 95% CI [0.99, 1.03], $p = 0.228$）和教育程度（$B = −0.01$, $Wald = 1.55$, $OR = 0.99$, 95% CI [0.97, 1.01], $p = 0.214$）对利他性惩罚均不显著；相较于男性被试，女性被试做出利他性惩罚的可能性显著更低（$B = −0.24$, $Wald = 15.61$, $OR = 0.79$, 95% CI [0.70, 0.89], $p < 0.001$），具体而言，在相似情境下女性被试做出利他性惩罚的可能性比男性被试低 21%；主观社会阶层对利他性惩罚具有显著的正向影响（$B = 0.07$,

Wald = 16.70, *OR* = 1.08, 95% CI [1.04, 1.11], $p < 0.001$），具体而言，被试的主观社会阶层每上升一个等级，做出利他性惩罚的可能就上升 8%。

接着，仍然以主观社会阶层为预测变量，反抗上司不端行为这一利他性惩罚为结果变量做二元 Logistic 回归（$N = 4864$），模型拟合度良好（$\chi^2(5, 4864) = 70.46$, $p < 0.001$; Hosmer-Lemeshow 检验不显著: $p = 0.056$）。回归分析结果显示：年龄（$B = 0.001$, *Wald* = 0.10, *OR* = 1.00, 95% CI [0.997, 1.005], $p = 0.747$）、收入（$B = 0.01$, *Wald* = 1.51, *OR* = 1.01, 95% CI [0.99, 1.03], $p = 0.219$）对利他性惩罚均不显著；教育程度（$B = 0.03$, *Wald* = 7.85, *OR* = 1.03, 95% CI [1.01, 1.06], $p = 0.005$）对利他性惩罚具有显著的正向影响，被试的学历每上升一个等级，做出利他性惩罚的可能就上升了 3%；女性被试做出利他性惩罚的可能性比男性低 30%（$B = -0.35$, *Wald* = 32.53, *OR* = 0.70, 95% CI [0.62, 0.79], $p < 0.001$）；主观社会阶层对利他性惩罚具有显著的正向影响（$B = 0.05$, *Wald* = 8.74, *OR* = 1.06, 95% CI [1.02, 1.09], $p = 0.003$），阶层每上升一个等级，被试做出利他性惩罚的可能就上升 6%。

研究 1 的结果支持了假设 1：在控制其他变量的情况下，主观社会阶层正向预测了被试的利他性惩罚，这一关系对举报社会问题和反抗上司不端行为这两种情境均成立。这一发现也有助于比较客观社会阶层和主观社会阶层对行为的影响。收入和教育程度通常被认为是反映个体客观社会阶层的两个重要指标（王艳丽等, 2017; Stephens et al., 2014），但在举报情境中这两个变量对利他性惩罚的影响均不显著，而在反抗上司情境中，仅教育的影响显著；而主观社会阶层在两种情境下均能较好地预测被试的惩罚行为，这从侧面回应了以往研究者的结论：客观社会阶层与主观社会阶层对个体行为的影响是相互独立的（Manstead, 2018），并且后者是比前者更好的预测指标（Cohen et al., 2008; Wolff et al., 2010）。

3 研究 2

研究 1 表明阶层正向影响了利他性惩罚，对于该现象的一种解释是不同阶层具有不同的成本敏感性（Paulsen & John, 2002）。如果这一解释成立，那么我们可以预测在高成本的惩罚中，阶层对惩罚的预测作用显著，而在低成本的惩罚中，这一预测作用可能就不复存在，因为在低成本条件下，不同阶层的成本敏感性差异可能无法得到充分体现。研究 2 根据 Molho 等（2020）的建议，将利他性惩罚区分为直接惩罚与间接惩罚，前者包括肢体对抗和言语指责，而后者包括流言（gossip）和社交回避（social avoidance）。虽然直接惩罚和间接惩罚所直观体现的是惩罚形式的差异，但两者之间的成本差异得到了许多研究者的认可（Archer &

Coyne, 2005; Balafoutas et al., 2014; Molho et al., 2020），即直接惩罚体现为高成本惩罚，而间接惩罚更像是一种低成本惩罚。根据上述区分，我们推测在直接惩罚中阶层的预测作用显著，而在间接惩罚中不显著。

3.1 被试

我们通过在线平台“Credamo 见数”向全国共投放了 500 份问卷，首先要求被试回忆过去一个月里所亲眼目睹的一件最严重的违规事件，并强调该事件对被试并没有发生直接影响，而是对其他人产生了负面影响。被试需要用 3~5 句话简要描述该违规事件，包括时间、地点和情节等，然后回答有关问题。要求报告描述违规事件是因为这样可以判断被试是否准确理解了利他性惩罚，从而避免将错误数据纳入分析。事实上，共有 50 人没有报告具体的违规事件或者所报告事件不属于利他性惩罚，剔除这 50 份数据后共得到有效问卷 450 份，有效回收率为 90.00%。被试平均年龄 $M = 30.21 \pm 5.73$ 岁，女性占 61.33%。

3.2 变量与研究工具

主观社会阶层 研究 2 中预测变量为主观社会阶层，测量同研究 1。

利他性惩罚 结果变量为利他性惩罚，在研究 2 中被区分为两种形式：直接惩罚和间接惩罚，各包括两个题项，回答“是”计分为 1，回答“否”计分为 0，两个题项得分相加即为直接惩罚和间接惩罚的最终分数。题项改编自 Molho 等（2020）的研究，具体如下：发生违规事件时，您事实上对违规者做出了何种反应？①与违规者发生肢体冲突、②朝违规者大声吼叫或与他/她争辩、③把违规者的不良行为告诉其他人和④尽量避免与违规者来往。前两者为直接惩罚的题项，后两者为间接惩罚的题项。

控制变量 控制变量包括被试的性别（1=男；2=女）、年龄、收入（1：月收入≤5000；2：5000<月收入≤10000；3：10000<月收入≤15000；4：月收入>15000）和教育程度（1=初中及以下；2=高中/中专/职高；3=本科/大专；4=研究生及以上）。

3.3 结果与讨论

我们首先检验了共同方法偏差问题。Harman 单因子法检验结果显示：未旋转得到的特征根大于 1 的因子共有 4 个，未旋转得到的第一个因子的变异量为 20.00%，远低于 40%的临界标准，且不同变量的测量方式相差甚远，因此可以认为不存在严重的共同方法偏差。各变量的描述统计和相关系数如表 2 所示。

表 2 研究 2 变量描述性统计与相关系数

变量	*M*	*SD*	1	2
1 主观社会阶层	5.45	1.24		
2 直接惩罚行为	0.34	0.54	0.19***	
3 间接惩罚行为	1.72	0.52	0.04	0.13**

注：$N = 450$，$^{**}p < 0.01$，$^{***}p < 0.001$。

两个结果变量均由两个二分题项相加而得，很难认为其属于连续变量，对其进行正态性检验，Kolmogorov-Smirnov 检验结果显示：无论是直接惩罚（$p < 0.001$）还是间接惩罚（$p < 0.001$）均不符合正态分布，因此统计方法我们采用了有序多分类 Logistic 回归。以直接惩罚为结果变量，主观社会阶层、年龄、性别、教育和收入为预测变量做有序多分类 Logistic 回归，平行线检验结果说明比例优势假设存在（$\chi^2(5, 450) = 4.22$, $p = 0.518$），满足开展有序多分类 Logistic 回归的前提条件；Deviance 拟合优度检验显示模型拟合良好（$\chi^2(539, 450) = 401.28$, $p = 1.000$）；模型拟合优度检验显示本模型优于只包含常数项的模型（$\chi^2(5, 450) = 29.97$, $p < 0.001$），说明本模型所纳入的预测变量中至少有一个对结果变量产生了显著影响。具体而言，收入（*Wald* = 0.12, *OR* = 1.06, 95% CI [0.77, 1.46], $p = 0.727$）和教育程度（*Wald* = 0.49, *OR* = 1.18, 95% CI [0.74, 1.89], $p = 0.484$）对直接惩罚的影响不显著；就性别而言，尽管单纯从数字上来看，男性做出惩罚的可能性更高，但差异并未达到显著性水平（*Wald* = 2.13, *OR* = 1.37, 95% CI [0.90, 2.08], $p = 0.144$）。与上述变量相反，年龄（*Wald* = 10.35, *OR* = 1.06, 95% CI [1.02, 1.10], $p = 0.001$）和主观社会阶层（*Wald* = 8.50, *OR* = 1.32, 95% CI [1.10, 1.59], $p = 0.004$）对直接惩罚的影响均显著，这意味着被试的年龄每增加一岁，对违规行为做出直接惩罚的可能性就上升了 6%；而阶层每上升一级，被试做出直接惩罚的可能性上升了 32%。

接着，以间接惩罚为结果变量，主观社会阶层、年龄、性别、教育和收入为预测变量做有序多分类 Logistic 回归，平行线检验结果说明比例优势假设存在（$\chi^2(5, 450) = 0.86$, $p = 0.973$），满足开展有序多分类 Logistic 回归的前提条件；Deviance 拟合优度检验显示模型拟合良好（$\chi^2(539, 450) = 417.05$, $p = 1.000$）；但模型拟合优度检验显示本模型和只包含常数项的模型无显著差异（$\chi^2(5, 450) = 6.91$, $p = 0.228$），说明本模型所纳入的预测变量对结果变量的影响均不显著，具体的分析也验证了这一结论。收入（*Wald* = 1.44, *OR* = 1.25, 95% CI [0.87, 1.78], $p = 0.230$），教育程度（*Wald* = 0.08, *OR* = 1.07, 95% CI [0.65, 1.78], $p = 0.781$）和性别

（*Wald* = 1.33, *OR* = 0.77, 95% CI [0.49, 1.20], $p = 0.248$）对间接惩罚影响不显著。和直接惩罚不同，以间接惩罚为结果变量时，年龄（*Wald* = 2.19, *OR* = 0.97, 95% CI [0.94, 1.01], $p = 0.139$）和主观社会阶层（*Wald* = 1.14, *OR* = 1.10, 95% CI [0.92, 1.33], $p = 0.286$）的影响也不显著。

有意思的是，在研究 1 中题项 D23（反抗上司不端行为）中同样包含着直接惩罚（当面抗议）和间接惩罚（暗地里报复），但在研究 1 中我们未能分开探讨阶层对两种惩罚形式的影响。和上述步骤一样，我们将当面抗议视为直接惩罚，而将暗地里报复视为间接惩罚，其余选项视为不惩罚，分别构造了 2 个虚拟变量。以社会阶层为预测变量，直接惩罚为结果变量进行二元 Logistic 回归，在控制性别、年龄、教育程度和年收入的情况下，阶层显著正向预测了直接惩罚（$B = 0.05$, *Wald* = 9.03, *OR* = 1.06, 95% CI [1.02, 1.09], $p = 0.003$）。以阶层为预测变量，间接惩罚为结果变量进行二元 Logistic 回归，在控制性别、年龄、教育程度和年收入的情况下，阶层对间接惩罚无显著影响（$B = -0.01$, *Wald* = 0.06, *OR* = 0.99, 95% CI [0.88, 1.10], $p = 0.809$）。这一发现是对研究 2 结果的有益补充，表明研究 2 基于真实生活事件调查的结果是稳健的。

上述结果大体上回应了研究 1 的发现：从总体上讲，包括收入和教育在内的客观社会阶层并不能很好地预测个体的惩罚行为，大部分时候其影响都不显著；相反，主观社会阶层大部分时候都是惩罚行为较好的预测指标。研究 2 进一步验证了我们的猜想，即惩罚成本在阶层影响惩罚行为的过程中确实发挥了一定的作用：阶层不同意味着成本敏感性不同（Paulsen & John, 2002），而在间接惩罚这种低成本条件下，这种成本敏感性的差异无法得到有效体现，因此阶层对惩罚的影响不显著。然而，研究 2 的局限也显而易见：第一，通过问卷所收集的数据虽然阐明了统计学意义上变量之间的关系，但我们仍然无法确定预测变量与结果变量之间是否真的存在因果关系；其次，区分直接惩罚和间接惩罚只能从侧面回答惩罚成本的影响，我们尚缺乏直接的证据；最后，研究 2 还无法回答成本在阶层影响惩罚的过程中具体发挥了怎样的作用，因此，我们设计了研究 3，通过直接操纵惩罚成本和社会阶层来回应上述问题。

4 研究 3

研究 2 通过区分两种惩罚形式证明惩罚成本在阶层影响利他性惩罚的过程中可能起到了调节作用，研究 3 通过直接操纵被试的主观社会阶层和惩罚成本为上述发现提供直接证据。需要说明的是，研究 3 通过引入潜在报复这种非支付形式的成本来操纵惩罚成本的高

低，原因有二：第一，有研究指出日常生活中报复是一种常见的惩罚成本（Engelmann & Nikiforakis, 2015）；其次，有研究进一步发现在惩罚成本数量相等的情况下，相比金钱支付，报复能更显著地抑制被试的惩罚行为（陈思静等, 2020）。

4.1 被试

使用软件 G*Power3.1 进行的功效分析（power analysis）显示：取中等效应量 $f = 0.25$，显著性水平 $\alpha = 0.05$，2 × 2 被试间方差分析至少需要 210 名被试才能达到 95%（$1 - \beta$）的统计检验力。实际参与研究 3 的被试为 240 名非心理学专业本科生。被试平均年龄为 20.60 ± 0.89 岁，女性占 40.52%。实验正式开始前，我们获得了所有被试的知情同意书。

4.2 设计与变量

研究 3 为 2（主观社会阶层：低/高）×2（惩罚成本：低/高）被试间设计。主观社会阶层的操纵基于 Kraus 等（2010）和 Piff 等（2010）的研究，将被试分成高阶层组和低阶层组；惩罚成本的操纵是通过引入受罚者的潜在报复将被试分成高成本组（有潜在报复）和低成本组（无潜在报复）。因变量是被试的惩罚行为，即被试在博弈任务中通过自行承担成本所扣减违规者的代币数。

4.3 程序

我们将被试随机分成四组（有 8 人未答对理解性检验题项，最终获得有效被试 232 人：高阶层高成本组 57 名，高阶层低成本组 57 名，低阶层高成本组 58 名，低阶层低成本组 60 名）。首先，根据 Kraus 等（2010）和 Piff 等（2010）的研究操纵了被试的主观社会阶层：我们先向被试展示了一幅包含 10 个层级的阶梯图，10 表示社会最顶层而 1 表示最底层；接着要求被试与位于阶梯最底层（高阶层组）/顶层（低阶层组）的人进行比较，这些人拥有最少/多的财富、接受了最低/高的教育、从事着最差/好的工作，特别是，我们要求被试考虑一下自己在收入、教育背景和工作方面与这些人有何不同；然后，我们要求被试想象一下自己和一个来自社会底层/顶层的人见面了，考虑到自己和这个人之间的阶层差距，自己会如何跟他/她打招呼？会和他/她说些什么？谈话会如何进行？用 3~5 个句子描述上述内容（不少于 20 字）；最后要求被试从 1~10 中选择一个整数来代表其认为自己所属的社会阶层。然后，被试观看 1 轮独裁者博弈，我们对博弈范式做了适当修改以突出分配者的违规程度：甲（分配者）和乙（接受者）通过共同努力获得了 50 代币的报酬，两人的贡献基本相同，但由甲来分配这些报酬，乙只能被动接受分配方案。最终甲选择分配给乙 10 代币，而将剩余的 40 代币都留给了自己。被试作为利益无关的第三方，共有 60 个代币，可以支付一定数量的代币来惩罚甲，规则为：每扣减甲 1 个代币，自己就需要支付 0.25 个代币，实验者要求被试

从 0~40 中选择一个数字来表示他/她所希望扣减甲的代币数。在高成本条件下，被试被告知他们的决策对甲可见，甲在知晓自己被惩罚后有可能报复被试，即甲可能反过来扣减被试同等的代币；在低成本条件下，被试被告知他们的决策对甲不可见，不会遭到甲的报复。被试做出惩罚决策后，回答以下问题：在上述情境中，如果您选择扣减甲 20 代币，您认为这个决策让您付出的成本如何（1=非常低；5=非常高）？被试回答上述问题后，实验者宣布实验结束，向被试解释实验目的并支付报酬。

4.4 结果与讨论

操纵有效性检验结果显示：和低阶层组相比，高阶层组的感知阶层显著更高（$t(230) = 7.01, p < 0.001, d = 0.92$, 95% CI [0.93, 1.65]）；和低成本组相比，高成本组的感知惩罚成本也显著更高（$t(230) = 2.64, p = 0.009, d = 0.34$, 95% CI [0.07, 0.52]），这意味着我们对社会阶层和惩罚成本的操纵是有效的。2×2 被试间方差分析表明，社会阶层的主效应显著（$F(1, 228) = 6.96, p = 0.009$, 偏 $\eta^2 = 0.03$），高阶层总体上更有可能惩罚违规者；惩罚成本的主效应显著（$F(1, 228) = 20.09, p < 0.001$, 偏 $\eta^2 = 0.08$），总体上被试在低成本条件下惩罚水平更高；且两者间的交互作用也显著（$F(1, 228) = 4.90, p = 0.028$, 偏 $\eta^2 = 0.02$）。简单效应分析表明（图 2）：一方面，在低成本条件下，高阶层被试（$M = 24.93, SE = 1.33$）和低阶层被试（$M = 24.37, SE = 1.30$）的惩罚行为无显著差异（$F(1, 228) = 0.09, p = 0.763$）；而在高成本条件下，相较于高阶层被试（$M = 21.93, SE = 1.33$），低阶层被试（$M = 15.52, SE = 1.32$）的惩罚行为显著偏低（$F(1, 228) = 11.67, p < 0.001$）。另一方面，对低阶层被试，高成本下的惩罚显著低于低成本下的惩罚（$F(1, 228) = 22.80, p < 0.001$）；但对高阶层被试，两种成本条件下的惩罚水平并无显著差异（$F(1, 228) = 2.53, p = 0.113$）。

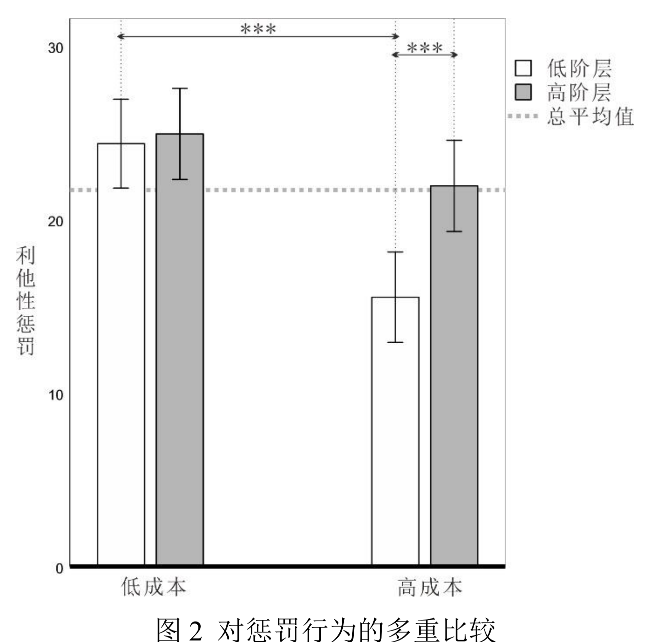


图 2 对惩罚行为的多重比较

注：图中误差线表示 95% CI, $^{***}p < 0.001$

Piff 等（2010）指出，诱导被试暂时体验较高或较低的社会阶层会激活相应的认知和动机，而这些认知与动机会产生相应的行为模式。本研究结果为上述观点提供了证据，并在一定程度上推进了对社会阶层影响惩罚行为的理解。当被试被诱导产生高阶层体验时，他们更有可能惩罚违规者，而这一效应在高成本条件下尤为明显，因为正如分析结果显示，在低成本条件下高低阶层的惩罚水平无显著差异，但在高成本条件下，高阶层的惩罚显著高于低阶层。这为假设 2 提供了明确的实验证据，同时也在一定程度上回应了研究 2 的结果。研究 2 表明，在成本较低的间接惩罚中，阶层无法有效地预测惩罚行为；而在成本较高的直接惩罚中，阶层可正向预测惩罚行为；而从本实验结果中也可以看到，总体来说，被试在高成本条件下的惩罚水平显著更低，这一效应对低阶层者尤为明显，因为低阶层者在高成本条件下的惩罚水平显著低于低成本条件，而高阶层者在这两种条件下的惩罚水平无显著差异。这表明不同阶层对成本的敏感性确实有所不同（Paulsen & John, 2002），也就是说，相较于高阶层者，是否存在报复对低阶层者的影响要大得多。

5 研究 4

研究 2 和研究 3 从成本角度考察了外部客观条件在阶层影响利他性惩罚中的影响，在最后一个研究中，我们从内部主观动机——公正世界信念——来分析利他性惩罚的心理机制，从而为本文主题提供一个相对完整的解释框架。

5.1 被试

使用软件 G*Power3.1 进行的功效分析显示：取中等效应量 $f^2 = 0.15$，显著性水平 $\alpha = 0.05$，多元回归分析至少需要 119 名被试才能达到 95%（$1-\beta$）的统计检验力。实际参与研究 4 的是 125 名社会被试，被试通过在线平台“Credamo 见数”招募。被试平均年龄 $M = 30.19 \pm 7.47$ 岁，女性占 62.40%。

5.2 变量与程序

主观社会阶层 预测变量为主观社会阶层，测量方法同研究 2。

公正世界信念 接着，我们测量了被试的公正世界信念，即研究 4 的中介变量。我们参考了 Wu 等（2011）的研究，通过 6 个题项（Cronbach's $\alpha = 0.75$）来测量被试的公正世界信念（例如，“从长远来说，我相信遭受不公正的人将会得到补偿”和“我确信公正总是可以战胜不公正”），6 个题项均为 6 点 Likert 量表：1＝ 完全不同意；6＝ 完全同意，6 个题项的平均得分越高表示被试的公正世界信念越强。

惩罚成本 调节变量为惩罚成本。在测量了被试的社会阶层和公正世界信念后，我们向被试展示了一个独裁者博弈，除惩罚成本外博弈范式和研究 3 基本相同。研究 3 通过操纵是否存在潜在报复将惩罚成本分为高低两个水平，在研究 4 中，我们告知被试他们的惩罚决策无论在何种条件下都对甲可见，并且有 80%的可能遭到甲（分配者）的报复，报复程度取决于被试的惩罚行为，共分 5 种情况，具体如下：您每扣减甲 1 代币，除了需要付出 0.25 个代币的成本外，还有可能遭受到来自甲 0.25/0.5/0.75/1/1.25 个代币的报复，此时您希望扣减甲多少代币？上述 5 种条件代表了不同的报复成本，即被试每付出 1 代币来惩罚违规者（也就是扣除违规者 4 代币），就有 80%的可能被违规者扣除 1/2/3/4/5 代币，在展示过程中按随机顺序出现。对惩罚成本采用了策略方法（strategy method）是因为它可以完整地揭示个体在不同条件下的反应（Jordan et al., 2016b），从而为分析提供更加丰富的数据。

利他性惩罚 结果变量为利他性惩罚，即在博弈中被试所希望扣减甲的代币数。

控制变量 控制变量包括性别、年龄、收入和教育程度，同研究 2。

5.3 结果与讨论

研究 4 中各变量的描述统计和相关系数如表 3 所示。

表 3 研究 4 变量描述性统计与相关系数

变量	*M*	*SD*	1	2	3	4	5	6
1 主观社会阶层	5.70	1.14						
2 公正世界信念	4.73	0.57	0.29^{**}					
3 惩罚-1	24.86	6.91	0.12	0.36^{***}				
4 惩罚-2	19.02	6.69	0.25^{**}	0.36^{***}	0.79^{***}			
5 惩罚-3	14.22	6.62	0.29^{***}	0.31^{***}	0.58^{***}	0.86^{***}		
6 惩罚-4	11.05	6.69	0.32^{***}	0.23^{**}	0.42^{***}	0.76^{***}	0.92^{***}	
7 惩罚-5	8.20	6.44	0.36^{***}	0.17	0.36^{***}	0.69^{***}	0.85^{***}	0.93^{***}

注：惩罚后的数字表示成本，$N = 125$，$^{**}p < 0.01$，$^{***}p < 0.001$。

研究 4 共有 125 名被试分别在 5 种不同的成本下，针对相同的违规行为实施了利他性惩罚，共 625 个数据点，包含个体和成本两个层面。其中惩罚和惩罚成本为成本层面的变量，公正世界信念、社会阶层和控制变量则为个体层面的变量。为验证图 1 的条件过程模型，我们以社会阶层为预测变量、惩罚为结果变量、公正世界信念为中介变量以及惩罚成本为调节变量，借助 Mplus8.3 使用极大似然估计检验了公正世界信念的跨层中介效应和惩罚成本的跨层调节效应。首先构建零模型以检验惩罚是否在个体层面拥有足够的变异，结果显示惩罚在个体层面的 ICC 为 0.29，这说明本研究需要使用多层线性模型来分析数据，以控制嵌套数据的非完全独立性所导致的误差（林琳, 2017）。其次，跨层中介模型的结果显示，中介变量公正世界信念对利他性惩罚的正向预测作用显著（$B = 2.42$, $SE = 0.76$, $p = 0.001$），且加入中介变量后社会阶层的预测作用仍然显著（$B = 1.05$, $SE = 0.47$, $p = 0.026$），这表明公正世界信念在社会阶层与利他性惩罚之间起着部分中介作用，这在一定程度上验证了假设 3。进一步检验惩罚成本的跨层调节作用，结果发现，公正世界信念和惩罚成本的交互项显著为负（$B = -0.90$, $SE = 0.27$, $p = 0.001$），而社会阶层和惩罚成本的交互项显著为正（$B = 0.44$, $SE = 0.14$, $p = 0.002$）。这意味着在社会阶层–利他性惩罚的直接路径中，随着成本的增加直接效应越来越强；而在社会阶层–公正世界信念–利他性惩罚的间接路径中，成本的上升逐渐减弱了间接效应。在上述跨层有调节的中介模型中，性别（$B = 0.38$, $SE = 1.14$, $p = 0.736$），年龄（$B = 0.06$, $SE = 0.11$, $p = 0.569$），收入（$B = -0.14$, $SE = 0.40$, $p = 0.732$）和教育程度（$B = 1.32$, $SE = 1.17$, $p = 0.256$）对惩罚的影响均不显著。

为了更完整地刻画惩罚成本的调节效应，我们参照沈伊默等(2017)的研究，使用 Edward 和 Lambert（2007）的路径分析技术，得到了在惩罚成本不同时（$M \pm 1SD$）社会阶层对利他性惩罚的直接效应、通过公正世界信念影响利他性惩罚的间接效应和总效应，结果如表 4 所

示。就直接效应而言，在成本低于均值一个标准差时，直接效应不显著，而在成本等于或大于均值时，直接效应的置信区间都不包含 0，此时直接效应显著且随成本的上升逐渐增强；就间接效应而言，在成本等于或小于均值一个标准差时，间接效应显著，但随着成本上升间接效应越来越小，当成本高于均值一个标准差时，间接效应不显著；而社会阶层影响利他性惩罚的总效应始终显著，随着惩罚成本的上升，总效应逐渐增强。上述分析结果表明，社会阶层既能直接影响利他性惩罚，又可以通过公正世界信念来间接地影响利他性惩罚，但这两种影响都是有条件的：如果惩罚成本非常高，那么阶层对利他性惩罚的间接影响就不显著了，而如果惩罚成本较低，阶层对利他性惩罚的直接影响就不显著。但从总体上来说，惩罚成本越高，社会阶层对利他性惩罚的影响就越大。综上所述，分析结果较好地支持了假设 3，并通过检验条件过程模型进一步细化了对惩罚成本和公正世界信念的分析。

表 4 惩罚成本调节社会阶层对利他性惩罚的直接效应、间接效应和总效应

惩罚成本	直接效应		间接效应		总效应	
	B	*SE*	*B*	*SE*	*B*	*SE*
$M - 1SD$	0.62	0.52	0.57^{**}	0.20	1.19^{*}	0.51
M	1.25^{**}	0.48	0.38^{**}	0.15	1.63^{**}	0.47
$M + 1SD$	1.88^{***}	0.52	0.19	0.13	2.07^{***}	0.52

注：$^{*}p < 0.05$，$^{**}p < 0.01$，$^{***}p < 0.001$。

6 总讨论

本文基于社会阶层检验了利他性惩罚在人群中的分布、心理机制和影响因素，一方面丰富了有关社会阶层的文献，将社会科学中两个重要的领域联结了起来；另一方面，本文也拓展了利他性惩罚的研究，并推进了对以下若干问题的理解。

6.1 为什么高阶层者的惩罚水平更高？

本文最重要的一个发现是社会阶层正向预测利他性惩罚，并且这一结论在四个研究中相当稳健。本文提供了两种视角来解释这一现象：基于惩罚成本的视角和基于公正世界信念的视角。我们首先探讨惩罚成本的影响。利他性惩罚是一种代价高昂的行为，所以也有研究者将其称为高成本惩罚（costly punishment）（Henrich et al., 2006; Rodrigues et al., 2020），这意味着行动成本必然从外部制约着惩罚行为。先前有大量研究表明成本越高，惩罚水平越低（陈世平, 薄欣, 2016; 范良聪等, 2013），这一点无论对高阶层者还是低阶层者均成立。不同的是，高阶层者拥有更多的资源，因此能在一定程度上抵消为惩罚所付出的代价，换言之，

高阶层者的成本敏感性更低（Paulsen & John, 2002）。因此，在控制其他因素的情况下，高阶层者更有条件做出利他性惩罚。研究 2 和 3 证实了上述观点：一方面，成本从总体上抑制了高低阶层者的惩罚水平，但随着成本的上升，高阶层者的下降幅度要低于低阶层者，这使得高阶层者表现出更高的惩罚水平。接着，我们讨论公正世界信念的影响。Hafer 和 Rubel（2015）指出，人们有内在的需求去维护公正世界信念，因为从长期来看，这有利于提升个体的主观幸福感和维护社会安定，人们有多种策略去维护公正世界信念，而利他性惩罚正是其中一种（Strelan et al., 2017）。Hafer 和 Rubel（2015）进一步指出，个体的公正世界信念水平越高，则越有可能采取措施来捍卫自己的信念。上述结论意味着个体的公正世界信念越高，则越有可能惩罚违规者。而从社会阶层的角度来看，高阶层者往往拥有更高的公正世界信念水平（Furnham & Procter, 1989），这意味着高阶层者不仅更有条件做出利他性惩罚，也更有意愿如此行动。这也就在一定程度上解释了为什么在本文的四个研究中，高阶层者总是表现出更高的惩罚倾向。

6.2 公正世界信念是利他性惩罚的另一种近因机制吗？

本文第二个重要发现是公正世界信念可能是社会阶层影响利他性惩罚的一种重要心理机制，这一发现为我们理解利他性惩罚的近因机制提出了新的思考方向。以往研究者多用愤怒作为利他性惩罚的近因解释（Fehr & Gächter, 2002; Fischer & Roseman, 2007），顺着这一方向，我们或许可以提出一种链式中介机制：高社会阶层–高公正世界信念水平–更容易愤怒–更多利他性惩罚，但目前的证据似乎指向了相反的方向，如 Dalbert（2002）和 Nesbit 等（2012）均发现，公正世界信念可以缓解人们的愤怒，即公正世界信念和愤怒之间更有可能存在负向关联，这一方面意味着上述链式中介机制可能并不成立，另一方面也暗示公正世界信念可能是一种独立于愤怒的心理机制，换言之，不需要愤怒，仅仅为了维护心中的某种信念就可以驱使人们做出利他性惩罚，即陈思静和杨莎莎（2020）所归纳的利他性惩罚的第一类动机。如果这一猜测成立，那么本文的发现为理解驱动利他性惩罚的心理机制提供了新的思路。当然，上述观点仅仅是基于本文发现和现有文献的一种推测，公正世界信念和愤怒在驱动利他性惩罚的过程中究竟是一种什么样的关系仍需更多研究。

6.3 公正世界信念为何只在惩罚成本低时发挥作用？

进一步，将公正世界信念和惩罚成本结合起来，本文提出了一个社会阶层影响利他性惩罚的条件过程模型，从这个模型中我们得到了本文的第三个新发现，即公正世界信念的中介作用是有边界条件的。具体而言，当惩罚成本较低时，阶层通过公正世界信念来影响利他性惩罚，但当成本变得非常高，这一中介作用就消失了；相反，阶层对利他性惩罚的直接作用

在惩罚成本很低的时候并不显著，只有当成本超过了一定的阈值，这一直接效应才会开始起作用。我们结合现有文献对这一现象提出了如下解释：不同阶层拥有不同的资源（Drentea, 2000; Oakes & Rossi, 2003），因此当惩罚成本较低时，不同阶层所拥有资源数量的差异并不会直接影响惩罚行为，因为此时的行动不会消耗太多资源，主要通过人们内心维护公正世界信念愿望的强弱来影响利他性惩罚；而随着客观成本的升高，内心信念的作用逐渐变弱，当成本变得很高时，内心的信念已经不足以支撑个体做出惩罚行为，这时影响惩罚的主要是人们所拥有资源的多少。换言之，当惩罚成本较低的时候，人们惩罚违规者更多的是为了维护内心中的公义信念，这个时候惩罚行为似乎更多地偏向非策略性考虑这一边；而当惩罚成本变得很高时，内心信念的驱动作用逐渐消失，影响惩罚与否更多的是人们手中资源的多少以及随之而来的成本敏感性，这个时候惩罚似乎更偏向基于成本-收益的策略性考虑。

6.4 主观阶层是否比客观阶层能更好地预测利他性惩罚？

本文共包含四个研究，除了研究 3 使用了较为同质化的大学生被试，在其他三个研究中均同时考察了客观社会阶层（收入和教育）与主观社会阶层对惩罚的影响。结果可总结如下：对于客观社会阶层，除了研究 1 中的教育程度对惩罚具有显著的正向影响外，其他情况下客观阶层的这两个指标都不显著；相反，对于主观社会阶层而言，四个研究表现出了稳定的结果，即主观阶层总是可以有效地预测利他性惩罚。这个发现提出了两个问题：第一，为什么主观阶层和客观阶层具有不同的作用？第二，为什么在不同的研究中客观阶层的作用也有所不同？对于第一个问题，我们推测一个重要原因是因为中国目前仍然处于一个快速变化的阶段（张跃等, 2020），而正如陈云松和范晓光（2016）指出，在处于快速变化期的社会中，客观阶层指标的作用会弱化，而主观社会阶层的作用会强化，本文的结果也回应了这一结论：本研究中基本上所有情况下，主观社会阶层的影响均大大高于客观社会阶层，这可能是由于中国目前特殊的历史时期造成的。此外，先前也有研究报告过类似的结果，如 Kraus 和 Tan（2015）的研究中，教育和收入对结果的影响表现出了不一致性，这似乎也在一定程度上说明客观阶层的影响不够稳定，总体表现不如主观阶层来得稳健（Cohen et al., 2008; Wolff et al., 2010）。总的说来，仅仅就利他性惩罚而言，主观阶层看起来确实是比客观阶层更好的预测指标。对于第二个问题，不一致的结果主要体现在教育这一指标在研究 1 和研究 2、4 中的差异，我们推测这可能是由于样本的差异所造成。研究 1 使用了 2013 年中国综合社会调查数据，而研究 2、4 是通过“见数”收集数据，而能接触到这一在线平台的被试可能本身就受过更好的教育，因此在教育水平上相对更为同质化，从而削弱了教育对结果变量的影响。

6.5 不足与未来研究方向

尽管取得了若干有意义的结果，但本文依然存在不足之处。首先，在研究 3 和 4 中我们尽管采用了一系列措施来尽可能地模拟报复，但受到研究条件的限制，这种报复最终仍然表现为金钱形式，即受罚者有可能反过来扣除惩罚者的报酬，而在现实生活中报复更可能表现为肢体冲突或言语攻击等，但这些形式受到研究伦理的限制很难在实验室环境中实现。未来研究可考虑如何在遵守研究伦理的情况下在实验室环境中更真实地模拟报复行为，如此可大大提高研究的生态效度。

其次，研究 1 使用了 2013 年中国综合社会调查数据来考察阶层与利他性惩罚之间的关系，被试主观社会阶层的均值是 4.33 ± 1.67，而研究 2 和研究 4 是通过在线平台收集数据，被试的阶层均值分别为 5.45 ± 1.24 和 5.70 ± 1.13，不仅阶层明显偏高，且离散程度更低，表明这两个研究的被试更为同质化。研究 3 没有测量自然状态下被试的主观社会阶层，但考虑到被试均为大学生，教育程度较高，因此可以合理地推测这些被试的阶层应该同样偏高。这一方面表明本文的结论具有较高的稳健性，因为在整体阶层偏低和偏高的样本中我们均得出了相似的结论；但另一方面也意味着，在后三个研究中得出的结论在推广时仍需谨慎，因为可能存在代表性不足的问题。

最后，目前尚无研究直接考察文化因素是如何影响社会阶层-利他性惩罚关系的，但有一些证据表明文化因素在上述关系中可能发挥了重要作用，如 Ding 等（2022）的最后通牒博弈实验显示，在美国被试中，主观认为自己更为贫穷的被试更有可能拒绝不公平分配；而在中国被试中，情况刚好相反。考虑到中国人在主观阶层分布方面和其他国家存在明显差异（陈云松, 范晓光, 2016），本文的发现是否能推广到其他文化中仍需进一步探索。

参考文献

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology*, *19*(6), 586–592.

Aharoni, E., Kleider-Offutt, H. M., Brosnan, S. F., & Watzek, J. (2019). Justice at any cost? The impact of cost-benefit salience on criminal punishment judgments. *Behavioral Sciences & The Law*, *37*(1), 38–60.

Anderson, C. M., & Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*, *54*(1), 1–24.

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, *9*(3), 212–230.

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 15924–15927.

Balliet, D., Mulder, L. B., & van Lange, P. A. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*(4), 594–615.

Bègue, L., & Bastounis, M. (2003). Two spheres of belief in justice: Extensive support for the bidimensional model of belief in a just world. *Journal of Personality*, *71*(3), 435–463.

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior*, *28*(3), 168–175.

Chen, S. J., Hu, H. M., & Yang, S. S. (2020). Payment vs. retaliation: Impact of cost form on third-party punishment. *Journal of Psychological Science*, *43*(2), 416–422.

[陈思静，胡华敏，杨莎莎. (2020). 支付与报复：成本形式对第三方惩罚的影响. *心理科学*, *43*(2), 416–422.]

Chen, S. J., & Yang, S. S. (2020). Motives of altruistic punishment. *Advances in Psychological Science*, *28*(11), 1901–1910

[陈思静，杨莎莎. (2020). 利他性惩罚的动机. *心理科学进展*, *28*(11), 1901–1910.]

Chen, S. P, & Bo, X. (2016). The influence of unfairness and punishment price to the demand of third-party punishment. *Studies of Psychology and Behavior*, (03), 372–376.

[陈世平，薄欣. (2016). 公平与惩罚价格对第三方惩罚需求的影响. *心理与行为研究*, (03), 372–376.]

Chen, Y. S., & Fan, X. G. (2016). Social class self-positioning, income inequality and subjective perceptions of mobility (2003—2013). *Social Sciences in China*, (12), 109–126.

[陈云松，范晓光. (2016). 阶层自我定位，收入不平等和主观流动感知 (2003—2013). *中国社会科学*, (12), 109–126.]

Christie, A. M., & Barling, J. (2009). Disentangling the indirect links between socioeconomic status and health: The dynamic roles of work stressors and personal control. *Journal of Applied Psychology*, *94*(6), 1466–1478.

Cohen, S., Alper, C. M., Doyle, W. J., Adler, N., Treanor, J. J., & Turner, R. B. (2008). Objective and subjective socioeconomic status and susceptibility to the common cold. *Health Psychology*, *27*(2), 268–274.

Côté, S., House, J., & Willer, R. (2015). High economic inequality leads higher-income individuals to be less generous. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), 15838–15843.

Dalbert, C. (2002). Beliefs in a just world as a buffer against anger. *Social Justice Research*, *15*(2), 123–145.

Ding, Y., Wu, J., Ji, T., Chen, X., & van Lange, P. A. (2017). The rich are easily offended by unfairness: Wealth triggers spiteful rejection of unfair offers. *Journal of Experimental Social Psychology*, *71*, 138–144.

Ding, Y., Wu, J., Ji, T., Chen, X., & van Lange, P. A. (2022). Perceptions of having less in the US but having more in China are associated with stronger inequality aversion. *Journal of Experimental Social Psychology*, *101*, 104342.

Drentea, P. (2000). Age, debt, and anxiety. *Journal of Health and Social Behavior*, *41*, 437–450.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*(1), 1–22.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, *35*, 553–572.

Engelmann, D., & Nikiforakis, N. (2015). In the long-run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare*, *45*(3), 561–577.

Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, *73*(6), 2017–2030.

Fan, L. C., Liu, L., & Liang, J. (2013). The demand for the third party punishment: An experimental examination. *Economic Research Journal*, (5), 98–111.

[范良聪, 刘璐, 梁捷. (2013). 第三方的惩罚需求: 一个实验研究. *经济研究*, (5), 98–111.]

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140.

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology*, *93*(1), 103–115.

Furnham, A., & Procter, E. (1989). Belief in a just world: Review and critique of the individual difference literature. *British Journal of Social Psychology*, *28*(4), 365–384.

Gowdy, J. M. (2004). Altruism, evolution, and welfare economics. *Journal of Economic Behavior & Organization*, *53*(1), 69–73.

Hafer, C. L., & Rubel, A. N. (2015). The why and how of defending belief in a just world. In J. M. Olson, & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 51, pp. 41–96). Burlington, WA: Academic Press.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770.

Hu, X. Y., Li, J., Lu, X. Z., & Guo, Y. Y. (2014). The psychological study of social class: Social cognitive perspective. *Journal of Psychological Science*, *37*(6), 1509–1517

[胡小勇, 李静, 芦学璋, 郭永玉. (2014). 社会阶层的心理学研究: 社会认知视角. *心理科学*, *37*(6), 1509–1517.]

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016a). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*(7591), 473–476.

Jordan, J. J., McAuliffe, K., & Rand, D. (2016b). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, *19*(4), 741–763.

Kraus, M. W., & Callaghan, B. (2016). Social class and prosocial behavior: The moderating role of public versus private contexts. *Social Psychological and Personality Science*, *7*(8), 769–777.

Kraus, M. W., Côté, S., & Keltner, D. (2010). Social class, contextualism, and empathic accuracy. *Psychological Science*, *21*(11), 1716–1723.

Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science*, *20*(1), 99–106.

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., & Keltner, D. (2012). Social class, solipsism, and contextualism: How the rich are different from the poor. *Psychological Review*, *119*(3), 546–572.

Kraus, M. W., & Tan, J. J. (2015). Americans overestimate social class mobility. *Journal of Experimental Social Psychology*, *58*, 101–111.

Kuang, Y., Wang, F., & Wang, Z. J. (2021). Social class and children's prosociality: A study in the context of China's dual urban-rural structure. *Social Psychological and Personality Science*, *12*(1), 63–70.

Law, A., & Law, J. (2006). Class and social stratification in contemporary Scotland. In G. Mooney, T. Sweeney, & A. Law. (Eds.), *Social care, health and welfare in contemporary Scotland* (pp. 69–98). Kynoch & Blaney.

Lerner, M. J. (1965). Evaluation of performance as a function of performer's reward and attractiveness. *Journal of Personality and Social Psychology*, *1*(4), 355–360.

Li, J., Cai, Q., Huang, L. H., Wang, N. E., & Zhang, Y. L. (2012). Neurobiological basis and cognitive mechanism of altruistic punishment. *Advances in Psychological Science*, *20*(5), 682–689.

[李佳, 蔡强, 黄禄华, 王念而, 张玉玲. (2012). 利他惩罚的认知机制和神经生物基础. *心理科学进展*, *20*(5), 682–689.]

Li, L. L. (2012). The stratification of social structure and the marketization of interests relations: New challenges facing social management in China. *Sociological Studies*, *27*(2), 1–19.

[李路路. (2012). 社会结构阶层化和利益关系市场化——中国社会管理面临的新挑战. *社会学研究*, *27*(2), 1–19.]

Lin, L. (2017). Integrating the theory of planned behavior and implementation intention to overcome procrastination. *Acta Psychologica Sinica*, *49*(07), 953–965.

[林琳. (2017). 拖延行为的干预: 计划行为理论和实施意向的影响. *心理学报*, *49*(07), 953–965.]

Manner, M., & Gowdy, J. (2010). The evolution of social and moral behavior: Evolutionary insights for public policy. *Ecological Economics*, *69*(4), 753–761.

Manstead, A. S. (2018). The psychology of social class: How socioeconomic status impacts thought, feelings, and behaviour. *British Journal of Social Psychology*, *57*(2), 267–291.

Molho, C., Tybur, J. M., Van Lange, P. A., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature Communications*, *11*, 3432.

National Survey Research Center at Renmin University of China. (2015). *Chinese general social survey (2013)*. Retrieved February 17, 2022, from http://cnsda.ruc.edu.cn.

[中国人民大学中国调查与数据中心. (2015). *中国综合社会调查(2013)*. 2022-02-17 取自 http://cnsda.ruc.edu.cn.]

Nelissen, R. M. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, *29*(4), 242–248.

Nesbit, S. M., Blankenship, K. L., & Murray, R. A. (2012). The influence of just-world beliefs on driving anger and aggressive driving intentions. *Aggressive Behavior*, *38*(5), 389–402.

Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: Current practice and steps toward a new approach. *Social Science & Medicine*, *56*(4), 769–784.

Paulsen, M. B., & John, E. P. S. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *The Journal of Higher Education*, *73*(2), 189–236.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, *56*, 1–28.

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784.

Piff, P. K., & Robinson, A. R. (2017). Social class and prosocial behavior: Current evidence, caveats, and questions. *Current Opinion in Psychology*, *18*, 6–10.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(11), 4086–4091.

Rodrigues, J., Liesner, M., Reutter, M., Mussel, P., & Hewig, J. (2020). It's costly punishment, not altruistic: Low midfrontal theta and state anger predict punishment. *Psychophysiology*, *57*(8), e13557.

Shen, Y. M., Zhou, W. J., Wei, L. H., & Zhang, Q. L. (2017). Benevolent leadership and subordinate innovative behavior: The mediating role of perceived insider status and the moderating role of leader-member exchange differentiation. *Acta Psychologica Sinica*, *49*(08), 1100–1112.

[沈伊默, 周婉茹, 魏丽华, 张庆林. (2017). 仁慈领导与员工创新行为: 内部人身份感知的中介作用和领导-部属交换关系差异化的调节作用. *心理学报*, *49*(08), 1100–1112.]

Stellar, J. E., Manzo, V. M., Kraus, M. W., & Keltner, D. (2012). Class and compassion: Socioeconomic factors predict responses to suffering. *Emotion*, *12*(3), 449–459.

Stephens, N. M., Fryberg, S. A., & Markus, H. R. (2011). When choice does not equal freedom: A sociocultural analysis of agency in working-class American contexts. *Social Psychological and Personality Science*, *2*(1), 33–41.

Stephens, N. M., Markus, H. R., & Phillips, L. T. (2014). Social class culture cycles: How three gateway contexts shape selves and fuel inequality. *Annual Review of Psychology*, *65*, 611–634.

Strelan, P., di Fiore, C., & van Prooijen, J. W. (2017). The empowering effect of punishment on forgiveness. *European Journal of Social Psychology*, *47*(4), 472–487.

Tanjitpiyanond, P., Jetten, J., & Peters, K. (2022). How economic inequality shapes social class stereotyping. *Journal of Experimental Social Psychology*, *98*, 104248.

Wang, Y. L., Guo, Y. Y., & Yang, S. L. (2017). Social class and the need for structure: A perspective from compensatory control theory. *Chinese Journal of Clinical Psychology*, *25*(2), 371–373.

[王艳丽, 郭永玉, 杨沈龙. (2017). 家庭社会阶层与其结构需求的关系: 补偿性控制理论视角. *中国临床心理学杂志*, *25*(2), 371–373.]

Wolff, L. S., Subramanian, S. V., Acevedo-Garcia, D., Weber, D., & Kawachi, I. (2010). Compared to whom? Subjective social status, self-rated health, and referent group sensitivity in a diverse US sample. *Social Science & Medicine*, *70*(12), 2019–2028.

Wei, Q. W., Li, M. Z., & Chen, X. C. (2018). Social class and social perception: Is warmth or competence more important? *Acta Psychologica Sinica*, *50*(2), 243–252.

[韦庆旺, 李木子, 陈晓晨. (2018). 社会阶层与社会知觉: 热情和能力哪个更重要? *心理学报*, *50*(2), 243–252.]

Wu, M. S., Yan, X., Zhou, C., Chen, Y., Li, J., Zhu, Z., ... & Han, B. (2011). General belief in a just world and resilience: Evidence from a collectivistic culture. *European Journal of Personality*, *25*(6), 431–442.

Yang, S. L., Rao, T. T., Yu, F., & Guo, Y. Y. (in press). Subjective class in psychology of social class: Contributions and deficiencies. *Advances in Psychological Science.*

[杨沈龙, 饶婷婷, 喻丰, 郭永玉. (印刷中). 主观阶层研究取向的贡献与弊端. *心理科学进展*.]

Yang, S. S., & Chen, S. J. (2022). Normative misperception in third-party punishment: An explanation from the perspective of belief in a just world. *Acta Psychologica Sinica*, *54*(3), 281–299.

[杨莎莎, 陈思静. (2022). 第三方惩罚中的规范错觉: 基于公正世界信念的解释. *心理学报*, *54*(3), 281–299.]

Zhang, Y., Guo, Y. Y., & Ding, Y. (2020). The effect of economic inequality on perception of class mobility. *Journal of Psychological Science*, (5), 1197–1203.

[张跃, 郭永玉, 丁毅. (2020). 经济不平等对阶层流动感知的影响及其机制. *心理科学*, (5), 1197–1203.]

Zheng, X. L., Xie, F. W., Ding, L., & Wang, X. (2021). Social class and college students' Internet altruistic behavior: Moderated mediating effect. *Psychological Development and Education*, *37*(2), 182–189.

[郑显亮, 谢方威, 丁亮,王雪. (2021). 社会阶层与大学生网络利他行为: 一个有调节的中介模型. *心理发展与教育*, *37*(2), 182–189.]

Subjective social class positively predicts altruistic punishment in economic games and real-life contexts

CHEN Sijing [1], YANG Shasha [2], WANG Hao [1], WAN Fenghua [1]

([1] *School of Economics and Management*, *Zhejiang University of Science and Technology*, *Hangzhou* 310023, *China*)

([2] *School of Economics*, *Shanghai University*, *Shanghai,* 200444, *China*)

Abstract

Altruistic punishment means that people privately bear the cost to punish norm violators, although the punishment yields no material gain. The positive effects of altruistic punishment on cooperation and norm maintenance are well documented and the possible mechanisms underlying these effects have also been widely tested. However, an important issue remains underexplored: Does people's social background influence their altruistic punitive behavior? If yes, how? This article uses four studies to test the relationship between altruistic punishment and social class, the psychological mechanisms underlying the relationship, as well as the boundary conditions.

Study 1 used the *Chinese general social survey (2013)* released by the National Survey Research Center at Renmin University of China to examine the relationship between altruistic punishment and social class. We selected two items as the dependent variables of Study 1 (D13: employees reported environmental pollution at their own cost; D23: employees retaliated against their foreign boss who insulted China). After screening the samples, a total of 4921 (for D13) and 4864 (for D23) valid data were obtained, respectively. Study 2 was a real-life event-based survey with 450 participants. In Study 2, we further investigated the relationship between social class and altruistic punishment under two forms (direct vs. indirect punishment). Study 3 was a 2 (social class: low/high) × 2 (punishment cost: low/high) between-participants design, and the main purpose was to demonstrate that punishment cost may play a moderating role in the process of how social class affects altruistic punishment. Based on the survey data, Study 4 proposed a conditional process model with belief in a just world as a mediating variable and punishment cost as a moderator variable, hereby providing a relatively complete explanatory framework for the impact of social class on altruistic punishment.

Study 1 showed that after controlling for educational level and annual income, participants' subjective social class could significantly positively predict their altruistic punishment. Study 2

demonstrated that the above results hold in direct punishment, but not in indirect punishment. The results of Study 3 showed that when the punishment cost increases, punitive behavior decreases overall, but the downward trend is more pronounced for lower-class participants. The results of Study 4 further demonstrated that social class affects altruistic punishment indirectly mainly through belief in a just world when punishment cost is low, whereas social class directly affects altruistic punishment when punishment cost is high.

To sum up, we have found evidence that high social class (vs. low social class) individuals are more willing to engage in altruistic punishment in economic games and real-life contexts, implying that in an increasingly stratified modern society, people's social background should not be ignored in the research of altruistic punishment. In addition, the results of this article also prove that on the one hand, altruistic punishment is at least partly a non-strategic sanction, because one force that drives people to punish is to protect their just belief, and on the other hand cost-benefit based considerations are not completely absent in altruistic punishment.

Keywords altruistic punishment, social class, belief in a just world, punishment cost